
MAX-STABLE MODELS FOR MULTIVARIATE EXTREMES

Authors: JOHAN SEGERS

– ISBA, Université catholique de Louvain,
Belgium (johan.segers@uclouvain.be)

Abstract:

- Multivariate extreme-value analysis is concerned with the extremes in a multivariate random sample, that is, points of which at least some components have exceptionally large values. Mathematical theory suggests the use of max-stable models for univariate and multivariate extremes. A comprehensive account is given of the various ways in which max-stable models are described. Furthermore, a construction device is proposed for generating parametric families of max-stable distributions. Although the device is not new, its role as a model generator seems not yet to have been fully exploited.

Key-Words:

- *Copula; domain of attraction; max-stable distribution; spectral measure; tail dependence.*

AMS Subject Classification:

- 60G70, 62G32

1. INTRODUCTION

Multivariate extreme-value analysis is concerned with the extremes in a multivariate random sample, that is, points of which at least some components have exceptionally large values. Isolating a single component brings us back to univariate extreme-value theory. In this paper, the focus will rather be on the dependence between extremes in different components. The issue of temporal dependence will be ignored, so that the dependence will be understood as cross-sectional only.

Mathematical theory suggests the use of max-stable models for univariate and multivariate extremes. The univariate margins must be one of the classical extreme-value distributions, Fréchet, Gumbel, and extreme-value Weibull, unified in the generalized extreme-value distributions. For the dependence structure, however, matters are more complicated. A complete characterization in the multivariate case was given in de Haan and Resnick (1977), describing extremal dependence in terms of spectral measures on a subset of the unit sphere. Statistically, this formulation is not always the most practical one, and a large number of other concepts have been proposed.

The aim of this paper is to provide a comprehensive account of the various ways in which max-stable models are described (Section 2). Second, a construction device is proposed for generating parametric families of max-stable distributions (Section 3). The device is not new as it appears already for instance in the theory of regularly varying multivariate time series in Basrak and Segers (2009) or in the concept of a D -norm in Falk et al. (2010). Still, its role as a model generator seems not yet to have been fully appreciated.

Inference on multivariate extremes via max-stable models for joint tails is an exciting field that is still in development. Inference methods can be non-parametric or parametric, and in the latter case, they can be likelihood-based, frequentist as well as Bayesian, or based on other techniques such as the method of moments or minimum distance estimation.

Max-stable models have the drawback that they are too coarse to describe tails of multivariate distributions with asymptotic independence sufficiently accurately. More refined models exist and they are the topic of an extensive literature, originating from the seminal paper of Ledford and Tawn (1996).

2. FUNDAMENTALS

Max-stable distributions arise out of the study of the limit behaviour of vectors of componentwise maxima. Their dependence structure can be described

via a number of dependence functions or via certain measures. All these objects are related and they highlight different aspects of the distribution of multivariate extremes.

2.1. Convergence of vectors of componentwise maxima

Let $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, for $i = 1, \dots, n$, be independent and identically distributed random vectors in dimension d . For each component $j = 1, \dots, d$, consider the sample maximum, and collect these maxima in a new random vector:

$$\mathbf{M}_n = (M_{n1}, \dots, M_{nd}), \quad M_{nj} = \max(X_{1j}, \dots, X_{nj}).$$

Observe that most of the time, the point \mathbf{M}_n does not belong to the sample, as the maxima in different components will typically occur at different time instances. Still, the analysis of the large-sample distribution of \mathbf{M}_n is a natural starting point for multivariate extreme-value theory.

Weak convergence of a sequence of random vectors implies weak convergence of each of the components. As in univariate extreme-value theory, it is therefore reasonable to apply increasing affine transformations to each of the margins and consider the sequence of random vectors

$$(2.1) \quad \left(\frac{M_{nj} - b_{nj}}{a_{nj}} : j = 1, \dots, d \right), \quad n = 1, 2, \dots$$

in terms of normalizing constants $a_{nj} > 0$ and b_{nj} . For each component $j = 1, \dots, d$, the weak limit of $(M_{nj} - b_{nj})/a_{nj}$ as $n \rightarrow \infty$ must be a univariate max-stable distribution, and necessary and sufficient conditions on the marginal distribution F_j of the j th component X_{ij} for such convergence to take place are well known, see for instance the monograph by de Haan and Ferreira (2006).

However, weak convergence of each of the d components in (2.1) is strictly weaker than joint convergence of the vector of normalized maxima. What is needed in addition is a condition on the dependence structure of the common joint distribution F of the random vectors \mathbf{X}_i . A convenient way to describe this dependence is via the copula C_1 of F , that is,

$$(2.2) \quad \Pr[\mathbf{X}_i \leq \mathbf{x}] = F(\mathbf{x}) = C_1(F_1(x_1), \dots, F_d(x_d)).$$

Assuming the margins F_1, \dots, F_d are continuous, as we will do henceforth, the copula C_1 of the distribution function F in (2.2) is unique and can be obtained as the joint distribution function of the random vector $(F_1(X_{i1}), \dots, F_d(X_{id}))$.

Weak convergence of a sequence of multivariate distribution functions to a limit with continuous margins is equivalent to weak convergence of the sequences of margins and of the sequence of copulas (Deheuvels, 1984; Galambos, 1987). The copula of the vector of component-wise maxima \mathbf{M}_n , and hence of any vector

that is obtained by applying increasing transformations to each of its components, is given by

$$(2.3) \quad C_n(\mathbf{u}) = \{C_1(u_1^{1/n}, \dots, u_d^{1/n})\}^n.$$

This can be checked from the fact that the joint distribution function of \mathbf{M}_n is F^n while its margins are F_j^n for $j = 1, \dots, d$. Hence, in order for the normalized maxima in (2.1) to converge in distribution to a nondegenerate limit, besides marginal convergence, the sequence of copulas C_n must converge as well.

The copulas that can arise as weak limits of C_n as $n \rightarrow \infty$ are called extreme-value copulas, that is, a copula C is called an extreme-value copula if there exists a copula C_1 such that, as $n \rightarrow \infty$,

$$(2.4) \quad \lim_{n \rightarrow \infty} \{C_1(u_1^{1/n}, \dots, u_d^{1/n})\}^n = C(u_1, \dots, u_d).$$

Extreme-value copulas arise as the class of possible limit copulas of vectors \mathbf{M}_n as $n \rightarrow \infty$. The copula C_1 is said to be in the domain of attraction of C . An extensive survey of the literature on extreme-value copulas is given in Gudendorf and Segers (2010).

The class of extreme-value copulas coincides with that of max-stable copulas, defined as follows. A copula C is max-stable if, for all $\mathbf{u} \in [0, 1]^d$ and $k = 1, 2, \dots$,

$$C(\mathbf{u}) = \{C(u_1^{1/k}, \dots, u_d^{1/k})\}^k.$$

In the setting of componentwise maxima of independent random samples, the previous identity means that the copula C_k of the random vector \mathbf{M}_k is the same for every sample size k . Clearly, a max-stable copula is also an extreme-value copula, being in its own domain of attraction. Conversely, each extreme-value copula can be shown to be max-stable: in (2.4), partition the sample of size $n = mk$ in m blocks of size k and let m tend to infinity for fixed k . Since the limit must not depend on k , the max-stability relation follows.

In summary, we have found that nondegenerate limit distributions of vectors of appropriately normalized componentwise maxima have extreme-value margins and an extreme-value or max-stable copula. Specifically, if

$$\Pr \left[\bigcap_{j=1}^d \left\{ \frac{M_{nj} - b_{nj}}{a_{nj}} \leq x_j \right\} \right] \xrightarrow{w} G(x_1, \dots, x_d), \quad n \rightarrow \infty,$$

then necessarily

$$G(x_1, \dots, x_d) = C(G_1(x_1), \dots, G_d(x_d))$$

with extreme-value margins G_1, \dots, G_d and an extreme-value copula C . Convergence of margins and convergence of copulas being two isolated issues, we can ignore the former and rather focus on the latter. In fact, the way in which the components are normalized is immaterial, as long as the transformations applied to the components are increasing.

2.2. Dependence functions

Take logarithms and apply a linear expansion to see that (2.4) is equivalent to

$$(2.5) \quad \lim_{n \rightarrow \infty} n \{1 - C_1(1 - n^{-1}x_1, \dots, 1 - n^{-1}x_d)\} \\ = -\log C(e^{-x_1}, \dots, e^{-x_d}) = \ell(\mathbf{x}), \quad \mathbf{x} \in [0, \infty)^d.$$

The limit ℓ is called the stable tail dependence function of C , going back to Huang (1992) and Drees and Huang (1998). The variable n tending to infinity along the positive integers can be replaced by a variable t tending to infinity along the positive reals.

The best known example is the Gumbel–Hougaard copula, for which $\ell_\theta(\mathbf{x}) = (x_1^\theta + \dots + x_d^\theta)^{1/\theta}$ in terms of a parameter $\theta \in [1, \infty]$ (Gumbel, 1960; Hougaard, 1986). The function ℓ_θ happens to be the θ -norm of the vector \mathbf{x} . The fact that ℓ_θ is a norm is no coincidence: in a remarkable paper by Molchanov (2008), a characterization is given of all the norms that can give rise to stable tail dependence functions. In Falk et al. (2010), $\ell(\mathbf{x})$ is called the D -norm of \mathbf{x} , with D referring to the Pickands dependence function, see below.

Let $\mathbf{X} = (X_1, \dots, X_d)$ denote a generic random vector in the original sample. The expression on the left-hand side in (2.5) contains the rescaled probability

$$(2.6) \quad 1 - C_1(1 - x_1/n, \dots, 1 - x_d/n) \\ = \Pr[F_1(X_1) > 1 - x_1/n \text{ or } \dots \text{ or } F_d(X_d) > 1 - x_d/n].$$

This probability concerns the event that at least one among the d components X_1, \dots, X_d should exceed a high percentile of its own distribution. The copula domain-of-attraction condition (2.4), originally involving the vector of componentwise sample maxima, has been replaced by a condition on the upper tail of a single random vector. This is akin to the familiar peaks-over-threshold approach in univariate extreme-value theory.

The tail copula, R , of Schmidt and Stadtmüller (2006) arises if all d components are required to exceed a large percentile simultaneously:

$$\lim_{n \rightarrow \infty} n \Pr[F_1(X_1) > 1 - x_1/n \text{ and } \dots \text{ and } F_d(X_d) > 1 - x_d/n] \\ = R(\mathbf{x}), \quad \mathbf{x} \in [0, \infty)^d.$$

Clearly, the relation between the functions ℓ and R is governed by the inclusion-exclusion formula. In higher dimensions, ℓ is somewhat more convenient than R , as setting some components x_j in the definition of ℓ to zero allows one to retrieve the lower-dimensional margins of the extreme-value copula. This is not possible for the tail copula R , as setting even a single x_j to zero immediately

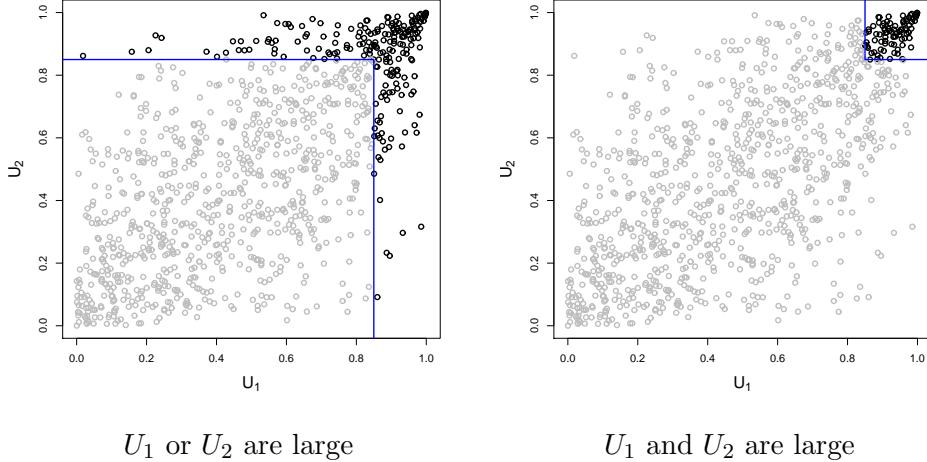


Figure 1: Left: the stable tail dependence function ℓ – at least one component should be large. Right: the tail copula R – all components must be large simultaneously. Here $U_j = F_j(X_j)$ for $j = 1, 2$.

yields $R(\mathbf{x}) = 0$. The difference between the two functions ℓ and R is depicted in Figure 1.

The dependence functions ℓ and R are homogeneous:

$$\begin{aligned}
 \ell(a \mathbf{x}) &= \lim_{t \rightarrow \infty} t \left\{ 1 - C_1 \left(1 - x_1/(t/a), \dots, 1 - x_d/(t/a) \right) \right\} \\
 &= \lim_{s \rightarrow \infty} a s \left\{ 1 - C_1 \left(1 - x_1/s, \dots, 1 - x_d/s \right) \right\} \\
 (2.7) \quad &= a \ell(\mathbf{x}), \quad a > 0, \mathbf{x} \in [0, \infty)^d,
 \end{aligned}$$

and similarly for R . It is therefore sufficient to consider the restriction of these functions to the unit simplex $\Delta_{d-1} = \{(w_1, \dots, w_d) \in [0, 1]^d : w_1 + \dots + w_d = 1\}$. The restriction of ℓ to Δ_{d-1} is called the Pickands dependence function, D , after Pickands (1981). By homogeneity,

$$\ell(\mathbf{x}) = (x_1 + \dots + x_d) D(w_1, \dots, w_d), \quad w_j = \frac{x_j}{x_1 + \dots + x_d}.$$

Frequently, the variable $w_d = 1 - w_1 - \dots - w_{d-1}$ is suppressed from the notation and D is written as a function of the arguments (w_1, \dots, w_{d-1}) only.

The probability on the right-hand side of (2.6) involves the union of the events $\{F_j(X_j) > 1 - x_j/n\}$, each of which has probability x_j/n , provided $0 \leq x_j \leq n$. As a consequence, we have the elementary bounds

$$\begin{aligned}
 &\max(x_1/n, \dots, x_d/n) \\
 &\leq \Pr[F_1(X_1) > 1 - x_1/n \quad \text{or} \quad \dots \quad \text{or} \quad F_d(X_d) > 1 - x_d/n] \\
 &\leq x_1/n + \dots + x_d/n.
 \end{aligned}$$

Multiplying by n and letting n tend to infinity, we obtain

$$(2.8) \quad \max(x_1, \dots, x_d) \leq \ell(x_1, \dots, x_d) \leq x_1 + \dots + x_d, \quad \mathbf{x} \in [0, \infty)^d.$$

By (2.5) and (2.8), an extreme-value copula C must satisfy

$$(2.9) \quad u_1 \cdots u_d \leq C(u_1, \dots, u_d) \leq \max(u_1, \dots, u_n), \quad \mathbf{u} \in [0, 1]^d.$$

The lower and upper bounds in the two previous displays can be attained, corresponding to the extreme cases of independence and perfect association. In particular, max-stable models are positive quadrant dependent. In fact, in Garralda-Guillem (2000), the stronger property is shown that bivariate extreme-value copulas are monotone regression dependent; see also Theorem 5.2.10 in Resnick (1987).

By (2.5), the copula C can be given in terms of the tail dependence function through

$$C(u_1, \dots, u_d) = \exp\{-\ell(-\log u_1, \dots, -\log u_d)\}, \quad \mathbf{u} \in (0, 1]^d.$$

In extreme-value theory, it is often convenient to standardize to other distributions than the uniform $(0, 1)$ law. The three most common forms are the unit Fréchet distribution, the Gumbel distribution, and the reverse exponential distribution, yielding respectively

$$(2.10) \quad C(e^{-1/x_1}, \dots, e^{-1/x_d}) = \exp\{-\ell(1/x_1, \dots, 1/x_d)\}, \quad \mathbf{x} \in (0, \infty)^d,$$

$$(2.11) \quad C(e^{-e^{-x_1}}, \dots, e^{-e^{-x_d}}) = \exp\{-\ell(e^{-x_1}, \dots, e^{-x_d})\}, \quad \mathbf{x} \in \mathbb{R}^d,$$

$$(2.12) \quad C(e^{x_1}, \dots, e^{x_d}) = \exp\{-\ell(-x_1, \dots, -x_d)\}, \quad \mathbf{x} \in (-\infty, 0)^d.$$

When using unit Fréchet margins, the notation $V(x_1, \dots, x_d) = \ell(1/x_1, \dots, 1/x_d)$ is often employed too.

2.3. The intensity measure

The transformation of the components X_j to uniform $(0, 1)$ random variables via the probability integral transform $F_j(X_j)$ has the disadvantage that all the action regarding the upper extremes is compressed to the neighbourhood of 1. Instead, for a univariate sequence ξ_1, ξ_2, \dots of independent and identically distributed random variables with common distribution function F_ξ , define the first exceedance time of the level x by

$$T(x) = \inf\{i = 1, 2, \dots : \xi_i > x\}.$$

If $F_\xi(x) < 1$, then $T(x)$ will be a geometric random variable with success probability equal to $1 - F_\xi(x)$. Its expectation,

$$\mathbb{E}[T(x)] = \frac{1}{1 - F_\xi(x)},$$

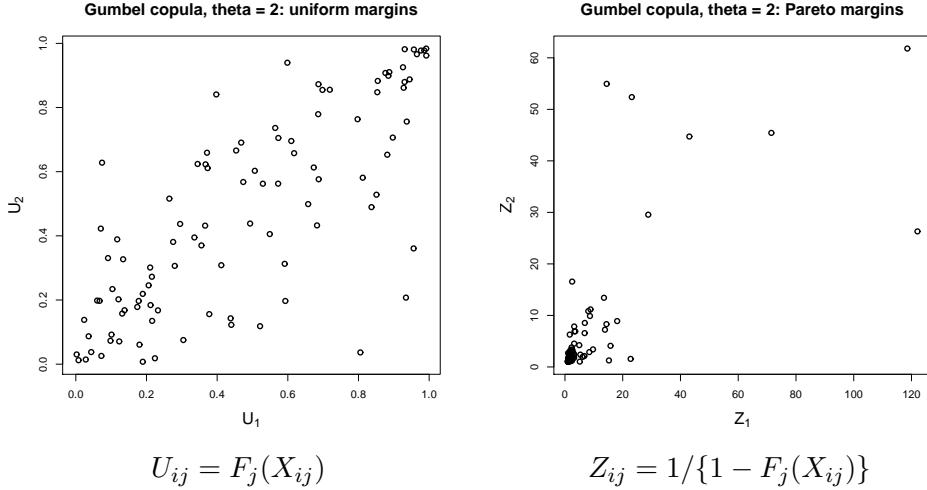


Figure 2: Two views on a sample: uniform (left) versus Pareto (right)

is called the return time of the level x .

Now let us apply this return time transformation to each of the d components of the random vector (X_1, \dots, X_d) . The return time of observation X_j is $Y_j = 1/\{1 - F_j(X_j)\}$. The law of Y_j is unit Pareto rather than uniform on $(0, 1)$, as $\Pr[Y_j > y] = \Pr[F_j(X_j) > 1 - 1/y] = 1/y$ for $y \geq 1$. We find that values of X_j corresponding to high percentiles of F_j are mapped to large values of Y_j . As is evident from Figure 2, extremes are magnified.

Now suppose that the copula C_1 is in the domain of attraction of an extreme-value copula with stable tail dependence function ℓ . Equation (2.5) says that the random vector $\mathbf{Y} = (Y_1, \dots, Y_d)$ satisfies

$$\begin{aligned} t \{1 - C_1(1 - x_1/t, \dots, 1 - x_d/t)\} &= t \Pr \left[\bigcup_{j=1}^d \{Y_j > t/x_j\} \right] \\ &= t \Pr \left[\mathbf{Y}/t \in ([0, \infty]^d \setminus [\mathbf{0}, 1/\mathbf{x}]) \right] \\ &\rightarrow \ell(\mathbf{x}), \quad t \rightarrow \infty. \end{aligned}$$

It follows that on the space $\mathbb{E}_d = [0, \infty]^d \setminus \{\mathbf{0}\}$, there exists a measure μ such that

$$(2.13) \quad \mathbb{E} \left[\sum_{i=1}^n I(\mathbf{Y}_i/n \in \cdot) \right] = n \Pr[\mathbf{Y}/n \in \cdot] \xrightarrow{v} \mu(\cdot), \quad n \rightarrow \infty.$$

The limit takes place in the mode of vague convergence of measures, meaning that $\lim_{n \rightarrow \infty} n \mathbb{E}[f(\mathbf{Y}/n)] = \int_{\mathbb{E}_d} f(\mathbf{x}) d\mu(\mathbf{x})$ for every bounded, continuous function f on \mathbb{E}_d that vanishes in a neighbourhood of the origin. Intuitively, when n grows large, the vector \mathbf{Y}/n is pulled towards the origin, in the neighbourhood of which the function f is zero. The intensity measure then only concerns the upper tail of

the distribution of \mathbf{Y} . The first expression in the previous display shows that μ expresses the limiting average intensity in space of the normalized sample cloud $\{\mathbf{Y}_1/n, \dots, \mathbf{Y}_n/n\}$. The stable tail dependence function acts as a distribution function for the intensity measure μ , as we have

$$\ell(\mathbf{x}) = \mu([0, \infty]^d \setminus [\mathbf{0}, 1/\mathbf{x}]).$$

2.4. Extreme profiles

Assume that the vector of return times $\mathbf{Y} = (Y_1, \dots, Y_d)$ with $Y_j = 1/\{1 - F_j(X_j)\}$ is large, that is, at least one of its components exceeds a high threshold. The relative sizes of the d components then inform us about the extremal dependence: are some components large simultaneously or is one specific component dominating all the other ones? Specifically, for $\mathbf{y} \in [0, \infty)^d \setminus \{0\}$ put

$$\begin{aligned} r(\mathbf{y}) &= y_1 + \dots + y_d \in (0, \infty), \\ \mathbf{w}(\mathbf{y}) &= (y_j/r(\mathbf{y}) : j = 1, \dots, d) \in \Delta_{d-1}, \end{aligned}$$

to be thought of as the magnitude and the profile of the vector \mathbf{y} , respectively. The spectral measure H lives on the unit simplex and is defined by

$$H(B) = \mu(\{\mathbf{y} : r(\mathbf{y}) > 1, \mathbf{w}(\mathbf{y}) \in B\}),$$

for Borel subsets B of Δ_{d-1} .

The measure convergence in (2.13) implies that

$$\begin{aligned} (2.14) \quad \mathbb{E} \left[\sum_{i=1}^n I\{r(\mathbf{Y}_i) > n, \mathbf{w}(\mathbf{Y}_i) \in \cdot\} \right] \\ = n \Pr[r(\mathbf{Y}) > n, \mathbf{w}(\mathbf{Y}) \in \cdot] \xrightarrow{w} H(\cdot), \quad n \rightarrow \infty, \end{aligned}$$

with vague convergence being replaced by weak convergence because the state space Δ_{d-1} is compact. In particular,

$$(2.15) \quad \Pr[\mathbf{w}(\mathbf{Y}) \in \cdot \mid r(\mathbf{Y}) > t] \xrightarrow{d} \frac{H(\cdot)}{H(\Delta_{d-1})} = Q(\cdot), \quad t \rightarrow \infty,$$

meaning that the conditional distribution of the profile $\mathbf{w}(\mathbf{Y})$ given that the magnitude $r(\mathbf{Y})$ is large converges in distribution to the normalized spectral measure Q .

The spectral measure H and the profile distribution $Q(\cdot)$ are alternative, equivalent ways of describing the extreme-value copula C . Indeed, homogeneity of ℓ in (2.7) implies homogeneity of μ :

$$(2.16) \quad \mu(a \cdot) = a^{-1} \mu(\cdot), \quad a > 0.$$

As a consequence, the intensity measure satisfies

$$(2.17) \quad \begin{aligned} \mu(\{\mathbf{y} : r(\mathbf{y}) > z, \mathbf{w}(\mathbf{y}) \in B\}) &= z^{-1} \mu(\{\mathbf{y} \mid r > 1, \mathbf{w} \in B\}) \\ &= z^{-1} H(B) \end{aligned}$$

for $z > 0$ and for Borel sets B of Δ_{d-1} . That is, when expressing a point \mathbf{y} in the coordinates (r, \mathbf{w}) , the intensity measure μ factorizes into a product measure on $(0, \infty) \times \Delta_{d-1}$ given by $r^{-2} dr H(d\mathbf{w})$. Equation (2.17) leads to

$$(2.18) \quad \int f(\mathbf{y}) \mu(d\mathbf{y}) = \int_{\Delta_{d-1}} \int_0^\infty f(r\mathbf{w}) r^{-2} dr H(d\mathbf{w})$$

for μ -integrable functions f , showing how to recover μ and thus ℓ and C from H . The special case where f is equal to the indicator function of the set $\{\mathbf{y} : \max_j x_j y_j > 1\}$ for some $\mathbf{x} \in [0, \infty)^d$ yields, after some computation,

$$(2.19) \quad \ell(\mathbf{x}) = \int_{\Delta_{d-1}} \max_{j=1, \dots, d} (w_j x_j) H(d\mathbf{w}).$$

Incidentally, this representation of ℓ implies that ℓ must be convex. By specializing the bounds in (2.8) to the unit vectors in \mathbb{R}^d , one finds that the spectral measure H must satisfy the constraints

$$(2.20) \quad 1 = \ell(\mathbf{e}_j) = \int_{\Delta_{d-1}} w_j H(d\mathbf{w}), \quad j = 1, \dots, d.$$

It follows that the total mass of H is equal to

$$H(\Delta_{d-1}) = \int (w_1 + \dots + w_d) H(d\mathbf{w}) = d.$$

Thanks to this property, it is possible to recover the spectral measure H from the profile distribution Q . From (2.20), it then follows that a random vector $\mathbf{W} = (W_1, \dots, W_d)$ on Δ_{d-1} with law equal to Q must satisfy

$$(2.21) \quad \mathbb{E}_Q[W_j] = \int_{\Delta_{d-1}} w_j Q(d\mathbf{w}) = 1/d, \quad j = 1, \dots, d.$$

In Section 3, we will see that any such law Q can appear as the profile distribution of a d -variate max-stable distribution.

In case of asymptotic independence, $\ell(\mathbf{x}) = x_1 + \dots + x_d$, the profile distribution Q is equal to the discrete uniform distribution on the d vertices of Δ_{d-1} : asymptotically, only one component can be large at a time. In the case of asymptotic perfect dependence, $\ell(\mathbf{x}) = \max(x_1, \dots, x_d)$, the profile distribution Q is degenerate at the center $(1/d, \dots, 1/d)$ of Δ_{d-1} : all components are equally large. These two extreme cases are depicted in Figure 3.

To show the ease with which coefficients related to extremal dependence can be computed, consider the random variable

$$N(t) = \sum_{j=1}^d I\{F_j(X_j) > 1 - 1/t\},$$

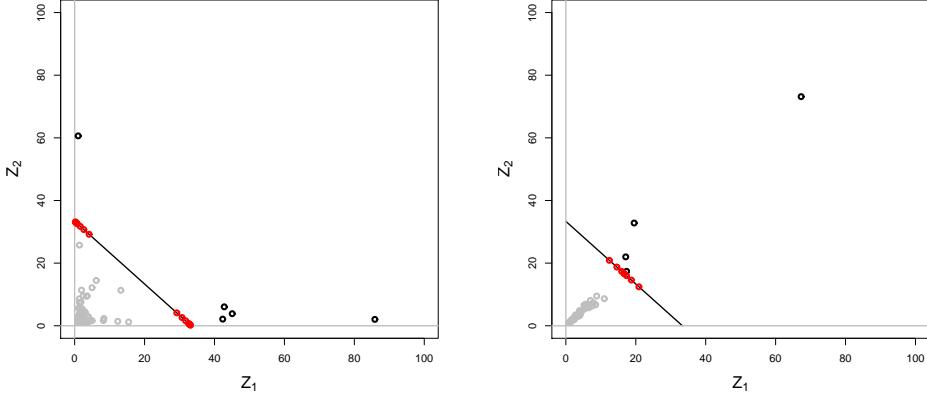


Figure 3: Asymptotic independence (left) versus asymptotic perfect dependence (right).

counting the number of components that exceed a high percentile. The following dependence coefficients have natural interpretations.

- *Trouble is in the air:*

$$\lim_{t \rightarrow \infty} t \Pr[N(t) \geq 1] = \ell(1, \dots, 1) = \int_{\Delta_{d-1}} \max(w_1, \dots, w_d) H(d\mathbf{w}).$$

- *Multiple failures:*

$$\lim_{t \rightarrow \infty} t \Pr[N(t) \geq k] = \int_{\Delta_{d-1}} w_{(d-k+1)} H(d\mathbf{w}),$$

for $k = 1, \dots, d$, where $w_{(1)} \leq \dots \leq w_{(d)}$ denote the order statistics of (w_1, \dots, w_d) .

- *The sky is falling:*

$$\lim_{t \rightarrow \infty} t \Pr[N(t) = d] = R(1, \dots, 1) = \int_{\Delta_{d-1}} \min(w_1, \dots, w_d) H(d\mathbf{w}).$$

- *System collapse – how bad will it get?*

$$\lim_{t \rightarrow \infty} \mathbb{E}[N(t) - k \mid N(t) \geq k] = \frac{\int (w_{(1)} + \dots + w_{(d-k)}) H(d\mathbf{w})}{\int w_{(d-k+1)} H(d\mathbf{w})},$$

for $k = 1, \dots, d-1$.

3. CONSTRUCTING MODELS

There is a fairly large number of parametric max-stable models available; see for instance the overviews in Kotz and Nadarajah (2000) and Beirlant et al. (2004). In a search for flexible models in large dimensions, new families are still being constructed, as in Ballani and Schlather (2011), Boldi and Davison (2007), Cooley et al. (2010) and Fougères et al. (2009). In this section, a simple construction device will be proposed and illustrated.

From Section 2, we recall that max-stable models for extremal dependence can be represented either via the extreme-value copula C , the stable tail dependence function ℓ , the intensity measure μ , or the spectral measure H and its normalized version, the profile distribution Q . However, as these objects must satisfy certain constraints, construction of parametric models is not obvious, particularly in high dimensions. Even if flexible parametric forms can be found, interpretation of the model parameters may not be obvious. In addition, when working with the spectral measure or profile distribution, the passage to lower-dimensional margins can be awkward, as the conditioning events in (2.15) will be different according to which components are selected.

3.1. A construction device

Let Z be a unit Fréchet random variable, that is, $\Pr(Z \leq z) = \exp(-1/z)$ for $z > 0$. Let $\mathbf{A} = (A_1, \dots, A_d)$ be a random vector, independent of Z , such that $0 < \mathbb{E}[\max(A_j, 0)] < \infty$ for every $j \in \{1, \dots, d\}$. Consider the random vector

$$(3.1) \quad \mathbf{X} = (X_1, \dots, X_d) = (A_1 Z, \dots, A_d Z).$$

The max-stable attractor of the distribution of \mathbf{X} can be explicitly calculated.

Lemma 3.1. *For $x = (x_1, \dots, x_d) \in (0, \infty)^d$, we have*

$$(3.2) \quad \begin{aligned} \lim_{n \rightarrow \infty} \Pr[X_1 \leq nx_1, \dots, X_d \leq nx_d]^n \\ = \exp\{-\mathbb{E}[\max(A_1/x_1, \dots, A_d/x_d, 0)]\}. \end{aligned}$$

Proof: Let $x \in (0, \infty)^d$. We have

$$\begin{aligned} \Pr[X_1 \leq x_1, \dots, X_d \leq x_d] &= \Pr[A_1 Z \leq x_1, \dots, A_d Z \leq x_d] \\ &= \Pr[A_1/x_1 \leq 1/Z, \dots, A_d/x_d \leq 1/Z] \\ &= \Pr[1/Z \geq \max(A_1/x_1, \dots, A_d/x_d)] \end{aligned}$$

The distribution of $1/Z$ is unit exponential. Since A and Z are independent,

$$\Pr[X_1 \leq x_1, \dots, X_d \leq x_d \mid A_1, \dots, A_d] = \exp\{-\max(A_1/x_1, \dots, A_d/x_d, 0)\}.$$

It follows that

$$\Pr[X_1 \leq x_1, \dots, X_d \leq x_d] = \mathbb{E}[\exp\{-\max(A_1/x_1, \dots, A_d/x_d, 0)\}].$$

Let $A^{(1)}, A^{(2)}, \dots$ be a sequence of independent and identically distributed copies of A . Fix positive integer n . We have

$$\begin{aligned} & \Pr[X_1 \leq nx_1, \dots, X_d \leq nx_d]^n \\ &= (\mathbb{E}[\exp\{-\max(A_1/x_1, \dots, A_d/x_d, 0)\}])^n \\ &= \mathbb{E}\left[\exp\left\{-\frac{1}{n} \sum_{i=1}^n \max(A_1^{(i)}/x_1, \dots, A_d^{(i)}/x_d, 0)\right\}\right]. \end{aligned}$$

Equation (3.2) now follows by the law of large numbers and the dominated convergence theorem. \square

The margins of the limit distribution function, say G , in (3.2) are equal to $G_j(x_j) = \exp\{-\mathbb{E}[\max(A_j, 0)]/x_j\}$ for $x_j > 0$. Assume that $\mathbb{E}[\max(A_j, 0)] = 1$ for all $j = 1, \dots, d$; this can always be achieved by rescaling the variables A_j if necessary. In that case, the margins of G are unit Fréchet. Comparing equations (3.2) and (2.10), we find that the stable tail dependence function of G is given by

$$(3.3) \quad \ell_{\mathbf{A}}(x_1, \dots, x_d) = \mathbb{E}[\max(x_1 A_1, \dots, x_d A_d, 0)]$$

The spectral measure H corresponding to ℓ in (3.3) can be identified too. Write $A_j^+ = \max(A_j, 0)$ and put $R = A_1^+ + \dots + A_d^+$. On the event $R > 0$, define $W_j = A_j^+/R$; on the event $R = 0$, the definition of W_j is immaterial – for definiteness, put $W_j = 1/d$ if $R = 0$. Note that \mathbf{W} takes values in Δ_{d-1} and that $\mathbb{E}[R] = \sum_{j=1}^d \mathbb{E}[A_j^+] = d$. We have $A_j^+ = RW_j$ and thus

$$\begin{aligned} \ell_{\mathbf{A}}(x_1, \dots, x_d) &= \mathbb{E}[R \max(W_1 x_1, \dots, W_d x_d)] \\ &= \mathbb{E}[\mathbb{E}[R \mid W_1, \dots, W_d] \max(W_1 x_1, \dots, W_d x_d)]. \end{aligned}$$

Comparing this expression with (2.19), we find that the spectral measure H of $\ell_{\mathbf{A}}$ is given by

$$(3.4) \quad H(d\mathbf{w}) = \mathbb{E}[R \mid \mathbf{W} = \mathbf{w}] \Pr[\mathbf{W} \in d\mathbf{w}],$$

that is, H is absolutely continuous with respect to the law of \mathbf{W} with Radon-Nikodym derivative equal to $\mathbb{E}[R \mid \mathbf{W} = \mathbf{w}]$. Similarly, the profile distribution Q satisfies

$$(3.5) \quad Q(d\mathbf{w}) = d^{-1} \mathbb{E}[R \mid \mathbf{W} = \mathbf{w}] \Pr[\mathbf{W} \in d\mathbf{w}],$$

Intuitively, this makes sense: profiles \mathbf{W} that on average yield larger values of R will have a larger contribution to the joint extremes of \mathbf{X} .

Incidentally, this construction shows that any probability distribution Q on Δ_{d-1} satisfying (2.21) can appear as the profile distribution of a d -variate max-stable distribution. Indeed, let the random vector \mathbf{W} on Δ_{d-1} have law Q and put $A_j = dW_j$ for all $j = 1, \dots, d$. As $A_1 + \dots + A_d = d$ by construction, the law of the random vector \mathbf{X} in (3.1) is in the domain of attraction of a d -variate max-stable distribution with profile distribution equal to Q .

If the dimension d is large, realistic models of extremal dependence should allow for the possibility that only some but not all components of a random vector are large simultaneously. In terms of the spectral measure or the profile distribution, this is encoded by the lower-dimensional faces of the unit simplex. For a non-empty subset I of $\{1, \dots, d\}$, let $\Delta_{d-1,I}$ denote the set of all \mathbf{w} in Δ_{d-1} such that $w_j > 0$ if $j \in I$ and $w_j = 0$ otherwise. If the probability of the event $\{\min_{j \in I} A_j > 0 \geq \max_{j \in I^c} A_j\}$ is non-zero, then by (3.4), the spectral measure and the profile distribution will put positive mass on $\Delta_{d-1,I}$. The set I contains the indices of the components that are large.

3.2. Examples

The simplicity of (3.3) is appealing and the reader is invited to apply the recipe in order to produce his or her own parametric models. In the remainder of the article, a number of well-known examples are worked out.

Example 3.1 (Independence). Suppose that $\Pr[A_j > 0 \geq \max(A_i : i \neq j)] = p_j$ for $p_j > 0$ and $p_1 + \dots + p_d = 1$. Then only component can be large at the time. After standardization ensuring that $\mathbb{E}[\max(A_j, 0)] = 1$ for all $j = 1, \dots, d$, we find $\ell_{\mathbf{A}}(x_1, \dots, x_d) = \mathbb{E}[\max(x_1 A_1, \dots, x_d A_d, 0)] = x_1 + \dots + x_d$, the stable tail dependence function of the independence copula.

Example 3.2 (Perfect dependence). Suppose that $A_j = a_j B$ with probability one for all $j = 1, \dots, d$, for some constants $a_j > 0$ and a random variable B such that $\mathbb{E}[\max(B, 0)]$ is positive and finite. Then the profile of an extreme vector is fixed. After standardization, $a_j = 1/\mathbb{E}[\max(B, 0)]$, the stable tail dependence function is that of perfect positive association, that is, $\ell_{\mathbf{A}}(x_1, \dots, x_d) = \max(x_1, \dots, x_d)$.

Example 3.3 (Discrete spectral measures). Suppose that the distribution of \mathbf{A} is discrete with a finite number of atoms. Specifically, suppose that $\Pr[\mathbf{A} = \mathbf{a}_k] = p_k$ for $k \in \{1, \dots, m\}$, where $\mathbf{a}_k \in \mathbb{R}^d$ and $p_k \in (0, 1)$ such that $\sum_{k=1}^m p_k = 1$. Via standardization, ensure that $1 = \mathbb{E}[\max(A_j, 0)] = \sum_{k=1}^m p_k a_{kj}^+$ for all $j = 1, \dots, d$. Put $r_k = a_{k1}^+ + \dots + a_{kd}^+$ and write $\mathbf{a}_k^+ = r_k \mathbf{w}_k$: if $r_k > 0$,

then $w_{kj} = a_{kj}^+ / r_k$, whereas if $r_k = 0$, then put $w_{kj} = 1/d$. It follows that

$$\begin{aligned}\ell_{\mathbf{A}}(x_1, \dots, x_d) &= \sum_{k=1}^m p_k \max(a_{k1}^+ x_1, \dots, a_{kd}^+ x_d) \\ &= \sum_{k=1}^m (p_k r_k) \max(w_{k1} x_1, \dots, w_{kd} x_d).\end{aligned}$$

We find that the spectral measure H and the profile distribution Q are discrete and are given by

$$H = \sum_{k=1}^m p_k r_k \delta_{\mathbf{w}_k}, \quad Q = \sum_{k=1}^m d^{-1} p_k r_k \delta_{\mathbf{w}_k},$$

with $\delta_{\mathbf{w}}$ a unit point mass at \mathbf{w} . The probabilities p_k are tilted with the magnitudes r_k , giving higher prominence to profiles \mathbf{w}_k that are associated to larger values of r_k .

Max-stable models with discrete spectral measures are called extreme-value factor models in Einmahl et al. (2011). Each of the m possible outcomes \mathbf{a}_k results in a different profile, according to the type of event or factor that triggered the extreme value.

Example 3.4 (Random indicators). Let

$$\ell(x, y) = \mathbb{E}[\max(xA, yB, 0)], \quad (x, y) \in [0, \infty)^2,$$

in terms of random variables A and B such that $\mathbb{E}[\max(A, 0)] = \mathbb{E}[\max(B, 0)] = 1$. Let (I, J) be a pair of random indicators, independent of the pair (A, B) , such that $\Pr[I = 1] = p$, $\Pr[J = 1] = q$, and $\Pr[I = J = 1] = r$. In the definition of ℓ , replace the pair (A, B) by the pair $(p^{-1}IA, q^{-1}JB)$; we assume that $0 < p \leq 1$ and $0 < q \leq 1$. The new stable tail dependence function is equal to

$$\begin{aligned}&\mathbb{E}[\max(p^{-1}xIA, q^{-1}yJB, 0)] \\ &= \mathbb{E}[\max(p^{-1}xA, q^{-1}yB)] \Pr[I = J = 1] \\ &\quad + \mathbb{E}[\max(p^{-1}xA, 0)] \Pr[I = 1, J = 0] + \mathbb{E}[\max(q^{-1}yB, 0)] \Pr[I = 0, J = 1] \\ &= \ell(p^{-1}x, q^{-1}y) r + p^{-1}x(p - r) + q^{-1}y(q - r)\end{aligned}$$

Writing $\alpha = r/p = \Pr[J = 1 \mid I = 1]$ and $\beta = r/q = \Pr[I = 1 \mid J = 1]$, we find

$$(3.6) \quad \ell_{\alpha, \beta}(x, y) = \ell(\alpha x, \beta y) + (1 - \alpha)x + (1 - \beta)y, \quad (x, y) \in [0, \infty)^2.$$

The new tail copula is simply

$$\begin{aligned}R_{\alpha, \beta}(x, y) &= x + y - \ell_{\alpha, \beta}(x, y) \\ &= \alpha x + \beta y - \ell(\alpha x, \beta y) = R(\alpha x, \beta y), \quad (x, y) \in [0, \infty)^2.\end{aligned}$$

This is an asymmetric, two-parameter extension of the original model. Imposing the equality constraint $\alpha = \beta = \theta \in [0, 1]$ yields the symmetric, one-parameter extension

$$(3.7) \quad \ell_\theta(x, y) = \theta \ell(x, y) + (1 - \theta)(x + y).$$

In higher dimensions, a vector of indicators (I_1, \dots, I_d) can serve to switch some components X_j ‘on’ or ‘off’. The dependence structure in these indicators then yields an extremal dependence structure for the random vector X . Specifically, let $p_j = \Pr[I_j = 1]$; we assume $0 < p_j \leq 1$. Let $\mathbf{A} = (A_1, \dots, A_d)$ be a random vector independent of (I_1, \dots, I_d) and such that $\mathbb{E}[\max(A_j, 0)] = 1$ for all $j = 1, \dots, d$. Then we can define a stable tail dependence function via

$$(3.8) \quad \begin{aligned} \ell_{\mathbf{p}}(x_1, \dots, x_d) &= \mathbb{E}[\max(p_1^{-1}x_1 I_1 A_1, \dots, p_d^{-1}x_d I_d A_d, 0)] \\ &= \sum_{\emptyset \neq c \subset \{1, \dots, d\}} p(c) \mathbb{E}[\max(p_j^{-1}x_j A_j : j \in c)] \end{aligned}$$

where $p(c) = \Pr[\{j = 1, \dots, d : I_j = 1\} = c]$. Keeping the law of \mathbf{A} fixed, the parameter vector \mathbf{p} is equal to a probability distribution $(p(c))_c$ on the non-empty subsets c of $\{1, \dots, d\}$.

In this way, hierarchical structures can easily be built. For instance, in dimension $d = 4$, we can think of $(I_1, I_2, I_3, I_4) = (J, J, K, K)$. One can also think of logit-type models for the indicators.

Example 3.5 (Marshall–Olkin copula). Applying the device in (3.6) to the function $\ell(x, y) = \max(x, y)$ yields the model

$$\ell_{\alpha, \beta}(x, y) = \max(\alpha x, \beta y) + (1 - \alpha)x + (1 - \beta)y = x + y - \min(\alpha x, \beta y).$$

The extreme-value copula associated to $\ell_{\alpha, \beta}$ is the Marshall–Olkin copula

$$C_{\alpha, \beta}(u, v) = uv \min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}), \quad (u, v) \in [0, 1]^2.$$

In higher dimensions, applying the device (3.8) to the function $\ell(\mathbf{x}) = \max(\mathbf{x})$, that is, $A_j = 1$ with probability one, we get the model

$$\ell_{\mathbf{p}}(x_1, \dots, x_d) = \sum_{\emptyset \neq c \subset \{1, \dots, d\}} p(c) \max(p_j^{-1}x_j : j \in c),$$

the spectral measure of which is discrete. Another stochastic interpretation of this model is provided in Embrechts et al. (2003).

Example 3.6 (Dirichlet model). Let $\alpha_1, \dots, \alpha_d$ be positive parameters and let Z_1, \dots, Z_d be independent $\text{Gamma}(\alpha_j, 1)$ random variables, that is, the density of Z_j is

$$f_j(z) = \frac{1}{\Gamma(\alpha_j)} z^{\alpha_j - 1} e^{-z}, \quad 0 < z < \infty.$$

Put $A_j = \alpha_j^{-1}Z_j$, a positive random variable with unit expectation. The stable tail dependence function associated with the random vector $\mathbf{A} = (A_1, \dots, A_d)$ is

$$\begin{aligned}\ell_{\mathbf{A}}(x_1, \dots, x_d) &= \mathbb{E}[\max(x_1A_1, \dots, x_dA_d)] \\ &= \mathbb{E}[\max(\alpha_1^{-1}x_1Z_1, \dots, \alpha_d^{-1}x_dZ_d)] \\ &= \mathbb{E}[(Z_1 + \dots + Z_d) \max(\alpha_1^{-1}x_1V_1, \dots, \alpha_d^{-1}x_dV_d)],\end{aligned}$$

where $V_j = Z_j/(Z_1 + \dots + Z_d)$. It is well known that the random vector (V_1, \dots, V_d) is independent of $Z_1 + \dots + Z_d$ and has a Dirichlet distribution with parameters $(\alpha_1, \dots, \alpha_d)$. We find

$$\begin{aligned}(3.9) \quad \ell_{\mathbf{A}}(\mathbf{x}) &= \mathbb{E}[(\alpha_1 + \dots + \alpha_d) \max(\alpha_1^{-1}x_1V_1, \dots, \alpha_d^{-1}x_dV_d)] \\ &= \frac{\Gamma(\sum_{j=1}^d \alpha_j + 1)}{\prod_{j=1}^d \Gamma(\alpha_j)} \int_{\Delta_{d-1}} \max_{j=1, \dots, d} (\alpha_j^{-1}x_jv_j) \prod_{j=1}^d v_j^{\alpha_j-1} dv_1 \dots dv_{d-1}.\end{aligned}$$

We recognize the Dirichlet model introduced in Coles and Tawn (1991).

Example 3.7 (Polynomial Pickands functions). In the Dirichlet model (3.9), put $d = 2$ and $\alpha_1 = \alpha_2 = 1$, to obtain

$$\ell(x, y) = 2 \int_0^1 \max\{xv, y(1-v)\} dv = x + y - \frac{xy}{x+y}.$$

The Pickands dependence function associated to ℓ is $D(t) = \ell(1-t, t) = 1-t(1-t)$ for $t \in [0, 1]$. Applying the transformation in (3.7) yields the mixture model

$$D(t) = 1 - \theta t(1-t), \quad t \in [0, 1],$$

in terms of a parameter $\theta \in [0, 1]$, introduced in Tawn (1988). This is the only model in which D is a quadratic polynomial. Applying the transformation in (3.6) yields the rational model

$$\begin{aligned}D_{\alpha, \beta}(t) &= \ell(\alpha(1-t), \beta t) + (1-\alpha)(1-t) + (1-\beta)t \\ &= 1 - \frac{\alpha\beta t(1-t)}{\alpha(1-t) + \beta t}, \quad t \in [0, 1],\end{aligned}$$

in terms of parameters $\alpha, \beta \in (0, 1]$.

More generally, bivariate max-stable models of which the Pickands dependence function is a higher-order polynomial can be obtained via the transformation in (3.7) applied to the function $\ell(x, y) = \mathbb{E}[\max(xA, yB)]$ when A and B are random sums

$$A = E_1 + \dots + E_J, \quad B = F_1 + \dots + F_K,$$

in terms of independent random variables $J, K, E_1, F_1, E_2, F_2, \dots$ such that J and K are counting random variables having finite support and unit expectation and where $E_1, F_1, E_2, F_2, \dots$ are unit exponential random variables. Polynomial Pickands dependence functions are studied in Guillotte and Perron (2012).

Example 3.8 (Schlather model). Let (S, T) be a bivariate normal random vector with standard normal margins and with correlation $\rho \in (-1, 1)$. Put $A = \sqrt{2\pi}S$ and $B = \sqrt{2\pi}T$, so that $E[\max(A, 0)] = E[\max(B, 0)] = 1$. The stable tail dependence function corresponding to the random pair (A, B) is

$$\begin{aligned}\ell_\rho(x, y) &= E[\max(xA, yB, 0)] \\ &= \int_{s=-\infty}^0 \int_{t=0}^\infty \sqrt{2\pi}yt \frac{1}{2\pi} e^{-(s^2+2\rho st+t^2)/2} ds dt \\ &\quad + \int_{s=0}^\infty \int_{t=-\infty}^0 \sqrt{2\pi}xs \frac{1}{2\pi} e^{-(s^2+2\rho st+t^2)/2} ds dt \\ &\quad + \int_0^\infty \int_0^\infty \sqrt{2\pi} \max(xs, yt) \frac{1}{2\pi} e^{-(s^2+2\rho st+t^2)/2} ds dt.\end{aligned}$$

After some tedious calculations, this can be shown to be

$$\ell_\rho(x, y) = \frac{1}{2}(x + y) \left(1 + \sqrt{1 - 2(\rho + 1) \frac{xy}{(x + y)^2}} \right),$$

see Schlather (2002). The Pickands dependence function corresponding to this model is

$$D_\rho(t) = \frac{1}{2} \left\{ 1 + \sqrt{1 - 2(\rho + 1)t(1-t)} \right\}, \quad t \in [0, 1].$$

Example 3.9 (Hüsler–Reiss model). Let (S, T) be a bivariate normal random vector with standard normal margins and with correlation $\rho \in (-1, 1)$. Let $\sigma > 0$ and put

$$A = \exp(\sigma S - \sigma^2/2), \quad B = \exp(\sigma T - \sigma^2/2).$$

The pair (A, B) is lognormal with unit expectations, and it yields the stable tail dependence function

$$\ell(x, y) = \int_{-\infty}^\infty \int_{-\infty}^\infty \max\{x e^{\sigma s}, y e^{\sigma t}\} e^{-\sigma^2/2} \frac{1}{2\pi} e^{-(s^2+2\rho st+t^2)/2} ds dt.$$

The double integral can be calculated in terms of the standard normal cumulative distribution function Φ , yielding

$$\ell_a(x, y) = x \Phi\left(\frac{a}{2} + \frac{1}{a} \log(x/y)\right) + y \Phi\left(\frac{a}{2} + \frac{1}{a} \log(y/x)\right)$$

with parameter $a = \sigma \sqrt{2(1 - \rho)}$. This is the model introduced in Hüsler and Reiss (1989).

ACKNOWLEDGMENTS

The author wishes to thank the journal Editors for the invitation to write this paper and for a detailed proofreading of an earlier version of the manuscript. The paper originates from a lecture given at the workshop on Copula Models and Dependence at the Centre de Recherches Mathématiques, June 6–9, 2011, in Montréal. Organizers and participants of the workshop are gratefully acknowledged for their support and their encouraging discussions, from which the paper has strongly benefited.

The author's research was supported by IAP research network grant No. P6/03 of the Belgian government (Belgian Science Policy) and by contract No. 07/12/002 of the Projet d'Actions de Recherche Concertées of the Communauté française de Belgique, granted by the Académie universitaire Louvain.

References

Ballani, F. and M. Schlather (2011). A construction principle for multivariate extreme value distributions. *Biometrika* 98, 633–645.

Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* 119, 1055–1080.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.

Boldi, M.-O. and A. C. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society. Series B* 69, 217–229.

Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B* 53, 377–392.

Cooley, D., R. Davis, and P. Naveau (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis* 101, 2103–2117.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York: Springer.

de Haan, L. and S. Resnick (1977). Limit theory for multidimensional sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 40, 317–337.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. In J. T. de Oliveira (Ed.), *Statistical Extremes and Applications*, pp. 117–130. Dordrecht: Reidel.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimates of the stable tail dependence function. *Journal of Multivariate Analysis* 64, 25–47.

Einmahl, J. H. J., A. Krajina, and J. Segers (2011). An M-estimator for tail dependence in arbitrary dimensions. Technical Report 2011/05, Université catholique de Louvain, ISBA. arXiv:1112.0905v1 [math.ST].

Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In S. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, Chapter 8, pp. 329–384. Elsevier.

Falk, M., J. Hüsler, and R.-D. Reiss (2010). *Laws of Small Numbers: Extremes and Rare Events* (3rd ed.). Basel: Birkhäuser.

Fougères, A.-L., J. P. Nolan, and H. Rootzén (2009). Models for dependent extremes using stable mixtures. *Scandinavian Journal of Statistics* 36, 42–59.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd ed.). Melbourne, FL: Robert E. Krieger Publishing Co. Inc.

Garralda-Guillem, A. I. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées. *C. R. Acad. Sci. Paris, Série I Math.* 330, 593–596.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In P. Jaworski, F. Durante, W. Härdle, and W. Rychlik (Eds.), *Copula theory and its applications* (Warsaw, 2009), Lecture Notes in Statistics – Proceedings, pp. 127–146. Berlin: Springer-Verlag.

Guillotte, S. and F. Perron (2012). Polynomial Pickands functions: Characterisation and inference. In preparation.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* 55, 698–707.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika* 73, 671–678.

Huang, X. (1992). *Statistics of bivariate extreme values*. Ph. D. thesis, Tinbergen Institute Research Series.

Hüsler, J. and R. Reiss (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters* 7, 283–286.

Kotz, S. and S. Nadarajah (2000). *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.

Molchanov, I. (2008). Convex geometry of max-stable distributions. *Extremes* 11, 235–259.

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* 5, 33–44.

Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* 33, 307–335.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* 75, 397–415.